SPECIAL BRIEFING

Artificial Intelligence: Oh Really? And Why Judges and Lawyers are Central to the Way we Live Now—But they Don't Know it

Stephen Mason*

Artificial intelligence; Autonomous vehicles; Causation; Design defects; Disclosure; Fault; Source codes

Something popularly called "artificial intelligence" has begun to infiltrate the realm of lawyers and policy-makers.[1] Professor Harel points out that "artificial intelligence" is a contradiction in terms, and refers to it as "algorithmic intelligence", a phrase that is to be preferred and is used in this article.[2] This article is in two parts. First, it aims to provide an explanation of algorithmic intelligence at a high level of generality; consider the meaning of algorithmic intelligence and some definitions; outline a number of tests devised to evaluate algorithmic intelligence; and consider criticism and successes.[3] The second part considers what the response of the judges, lawyers and legal academics ought be in the world in which we live now—the age of software code—colloquially called "algorithmic intelligence'.

Algorithmic intelligence

The meaning of algorithmic intelligence

It would be useful to determine what we mean by intelligence, but there is no general agreement about what is meant by the term. A useful alternative method is to consider the intellectual capacity of humans by considering a number of *characteristics* (this list is not exhaustive or in any order): learning or self-organisation in the broadest sense[4]; a capacity for logic (although formal logic can be unreasonable); abstract thought; the ability to recognise patterns; to comprehend complex ideas ("comprehend" naturally implies intelligence); to reason; to understand; carry out complex planning; to solve problems (other life forms solve problems); creativity (which might be as difficult to define as intelligence itself); self-awareness; perception (many animals perceive in sophisticated ways, such as to identify and select a "good" mate); communicate, a characteristic shared by all sentient beings and plants, although human language has a flexibility and open-endedness that no known animal communication system has (we are undoubtedly missing the sophistication of whale communications, and it appears that the vocabulary of prairie dogs is more advanced than any other animal language that has been decoded), and emotion and memory (characteristics shared by many sentient beings).

Commentators also include consciousness and empathy in this list. Regarding consciousness, Professor Häggström suggests, as do others, that we still do not know what this is.[5] Professor Russell and Peter Norvig describe consciousness as understanding, self-awareness, subject experience; they comment that:

> "Turing himself concedes that the question of consciousness is a difficult one, but denies that it has much relevance to the practice of AI … We agree with Turing – we are interested in creating programs that behave intelligently. The additional

* Stephen Mason is a barrister. The author thanks Nicholas Bohm, Derek Partridge, Professor Emeritus, University of Exeter, Jonathan Polnay of 5 King's Bench Walk and Martyn Thomas, CBE, FREng, Livery Company Professor of Information Technology, Gresham College and Director, UK Health and Safety Executive for their reviews of various iterations of this article. I am indebted to them for their comments and observations, all of which have improved the content. Suffice to say I am solely responsible for the content.

[1] The following list is far from exhaustive: Bert-Jaap Koops, Mireille Hildebrandt and David-Olivier Jaquet-Chiffelle, "Bridging the Accountability Gap: Rights for New Entities in the Information Society?" (2010) 11(2) *Minnesota Journal of Law, Science & Technology* 497; Nathalie Nevejans, *European Civil Law Rules in Robotics*, Study commissioned by the Legal Affairs Committee, European Parliament, PE 571.379 (European Union, 2016); *Civil Law Rules on Robotics*, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), P8_TA-PROV(2017)0051; Ryan Calo, A. Michael Froomkin and Ian Kerr (eds), *Robot Law* (Cheltenham: Edward Elgar Publishing, 2016); Chris Holder, Vikram Khurana, Faye Harrison and Louisa Jacobs, "Robotics and law: Key legal and regulatory implications of the robotics age (Part I)" (2016) 32(3) *Computer Law & Security Review* 383; Chris Holder, Vikram Khurana, Faye Harrison and Louisa Jacobs, "Robotics and law: Key legal and regulatory implications of the robotics age (Part II) (2016) 32(4) *Computer Law & Security Review* 557.

[2] David Harel, *Computers Ltd: What They Really Can't Do* (Oxford: Oxford University Press, 2000), p.194. In March 2012 Professor Harel wrote a new preface to the reprint, indicating that the underlying points made in his text remained: *http://www.wisdom.weizmann.ac.il/~harel/ltd.html#turingPreface* [Accessed 29 September 2017].

[3] This article does not consider the wider issue of how software code in use in making decision is opaque, unregulated, and when decisions are wrong, for which see Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown Publishing, 2016); for case law from Canada regarding the reliability of software code in actuarial tests, see *Ewart v R.* 2015 FC 1093, overturned on appeal in *R. v Ewart* 2016 FCA 203.

[4] Derek Partridge, *What Makes You Clever: The Puzzle of Intelligence* (Singapore: World Scientific, 2014), p.189, "non-adaptive intelligence is an oxymoron".

[5] Olle Häggström, *Here Be Dragons: Science, Technology and the Future of Humanity* (Oxford: Oxford University Press, 2016), pp.67–68 suggests, as do others, that we still do not know what this is; Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 3rd edn (Harlow: Pearson, 2016).

> project of making them conscious is not one that we are equipped to take on, nor one whose success we would wish to determine."[6]

Concerning empathy, arguably nobody is capable of truly understanding or feeling what another person is feeling.[7] Most of us can imagine what somebody else is feeling, but it will always be from our perspective, although the research of Dr Sarah Garfinkel, a Research Fellow (Neuroscience and Imaging) at the University of Sussex, indicates that some individuals have a physical ability to know what others are experiencing.[8] Software programs have been written for the purpose of interpreting our emotional state. Even though the software program is written to identify and interpret data, it is possible that the software code has been devised to lie.[9] Most, perhaps all, of these characteristics are indicative of intelligence only by virtue of the level manifest in humans rather than in other animals, although many of these suggested characteristics are found to a lesser degree in animals.[10]

We think, but nobody knows how we think. This raises the question posed by Professor Dreyfus in relation to the attempt to replicate the impossible[11]:

> "Why, in the light of these difficulties, do those pursuing Cognitive Simulation assume that the information processes of a computer reveal the hidden information processes of a human being, and why do those working in Artificial Intelligence assume that there must be a digital way of performing human tasks? To my knowledge, no one in the field seems to have asked himself these questions. In fact, artificial intelligence is the least self-critical field on the scientific scene. There must be a reason why these intelligent men almost unanimously minimize or fail to recognize their difficulties, and continue dogmatically to assert their faith in progress."

In the same context, Jaron Lanier makes a similar point[12]:

> "The antihuman approach to computation is one of the most baseless ideas in human history. A computer isn't even there unless a person experiences it. There will be a warm mass of patterned silicon with electricity coursing through it, but the bits don't mean anything without a cultured person to interpret them.
>
> This is not solipsism. You can believe that your mind makes up the world, but a bullet will still kill you. A virtual bullet, however, doesn't even exist unless there is a person to recognize it as a representation of a bullet. Guns are real in a way that computers are not."

Professor Johnston, in discussing his emergence of mind hypothesis, sums up the reality when discussing software code by pointing out the obvious[13]:

> "[C]omputers don't really think, they merely simulate thoughts, and although they certainly don't feel, they can be used to simulate the functions of feelings.
>
> …
>
> Meaning is not an inherent property of algorithms, because algorithms describe processes and not the content of those processes. An algorithm, for example, could possibly describe the process of seeing or feeling but not the qualitative nature of the conscious experience: what is actually seen (redness) or what is consciously felt (pain) … Meaning requires conscious experiences, which are the evolved emergent properties of biological brains.
>
> …
>
> Human conscious experiences are inherent emergent properties that arise from the complex arrangements and interactions between nerve cells, not transistors, so it is highly unlikely that any silicon-based machine could ever generate similar conscious attributes.
>
> …
>
> [C]onsciousness … is a property of biological tissue, and conscious experiences do not exist anywhere – not in rocks, nor plants, nor computers – outside the brains of living animals … All conscious experiences are emergent properties of biological brains, and they do not exist outside of those brains.
>
> …
>
> A computer does not compute anything unless a human operator assigns a meaning to the different states of the system; meaning is not an inherent property of algorithms. Unless somebody assigns a meaning to these symbolic states, then no computation is possible."

[6] Russell and Norvig, *Artificial Intelligence: A Modern Approach* (2016), p.1033; A.M. Turing, "Computing Machinery and Intelligence" (1950) 59(236) *Mind* 433.
[7] On empathy generally, see Jeremy Rifkin, *The Empathic Civilization: The Race to Global Consciousness in a World in Crisis* (New York: Penguin/Tarcher, 2009); Frans de Waal, *The Age of Empathy: Nature's Lessons for a Kinder Society* (New York: Souvenir Press, 2009); Simon Baron-Cohen, *Zero Degrees of Empathy: A New Theory of Human Cruelty* (London: Allen Lane, 2011).
[8] Sarah Garfinkel, "The Uncommon Senses: Introception", broadcast on BBC Radio 4 on 21 March 2017.
[9] Note the VW debacle, in which it transpired that the software code in the motor vehicles the company manufactured was designed to deceive, for which see *In re Volkswagen "Clean Diesel" Marketing, Sales Practices, and Products Liability Litigation*, 25 October 2016, Case No.3:15-md-2672 before Breyer J of the US District Court Northern District of California San Francisco Division (relates to: *FTC v Volkswagen Group of America Inc*, No.3:16-cv-1534 (N.D. Cal)): *Volkswagen Clean Air Act Civil Settlement*, United States Environmental Protections Agency, *https://www.epa.gov/enforcement/volkswagen-clean-air-act-civil-settlement*; Volkswagen AG entered pleas of guilty before U.S. District Court Judge Sean Cox in Detroit on 19 March 2017: for relevant documents, see *https://www.justice.gov/opa/pr/volkswagen-ag-agrees-plead-guilty-and-pay-43-billion-criminal-and-civil-penalties-six* [Both accessed 29 September 2017].
[10] This list is merely representative: Peter Godfrey-Smith, *Other Minds: The Octopus and the Evolution of Intelligent Life* (New York: William Collins, 2017); Frans de Waal, *Are We Smart Enough to Know How Smart Animals Are?* (London: Granta, 2016); Hal Whitehead and Luke Rendell, *The Cultural Life of Whales and Dolphins* (Chicago: University of Chicago Press, 2014).
[11] Hubert Dreyfus, *What Computers Still Can't Do: A Critique of Artificial Reason* (New York: Harper & Row, 1992), p.151.
[12] Jaron Lanier, *You Are Not A Gadget* (London/New York: Penguin, 2011), pp.26–27.
[13] Victor S. Johnston, *Why We Feel The Science of Human Emotions* (London: Perseus Books, 1999) pp.56, 58, 59, 182, 200, fn3.

It is possible to conclude that it depends on what you mean by intelligence as to how you define it. But does it matter that we cannot define intelligence? Are the debates about intelligence and consciousness relevant? One might ask why people insist on trying to establish whether inert machinery and software code written by humans can "think" or have "consciousness". The late John McCarthy, one of the founders of the concept of algorithmic intelligence, concluded that how the brain operates is not relevant for algorithmic intelligence—that is, if we think only in terms of human intelligence.[14] This is because the entire discussion of what intelligence means is irrelevant when discussing a machine made by humans, comprising software code written by human beings, and various items of hardware, all mined, engineered and manufactured by human beings, run by electricity which is in turn created by human beings.

There is another reason for asking whether this discussion is not as helpful as it might seem. It is axiomatic that we cannot independently verify that another person is thinking. We assume that they are thinking, otherwise communication becomes meaningless. One could argue that the same criteria ought to apply to software code. If we did apply the same criteria to software code, it is quickly obvious that communicating with software code is meaningless, other than for a highly restricted and specialist area that can be taken in logical steps, such as renewing motor vehicle tax. The logic in treating a response to a communication by software code in the same way as a communication with another human being is impeccable. This misses the point. Professor Turkle warns about how machines controlled by software code are used to interact with children at school. Depending on how such machines are used, there is a risk that a child will create a false relationship with the machine. Professor Turkle has said:

> "Why are we working so hard to set up a relationship that can only be 'as if'? The robot can never be in an authentic relationship. Why would we normalise what is false and in the realm of [a] pretend relationship from the start?"[15]

A human being is a biological entity with DNA and millions of years of evolutionary development, as well as personal history. Indeed, Ulric Neisser pointed out in 1963 three fundamental and interrelated characteristics of human thought[16]:

> "1) human thinking takes place in and contributes to a cumulative process of growth and development;
> 2) human thinking begins in an intimate association with emotions and feelings which is never entirely lost;
> 3) almost all human activity, including thinking, serves not one but a multiplicity of motives at the same time."

Humans use their knowledge, understanding of the world and physical experiences of sensations, such as sight, smell, taste, hearing and touch, to respond to a communication that is not explicitly pre-programmed and therefore pre-conceived, however limiting they may be as an individual.[17] The fact is, that software code replicates, as pointed out by Stephen Baker in relation to the IBM machine named Watson: "It comes up with a simulation of some kind of psychology and knowledge of humans. It doesn't understand anything."[18]

Some definitions of algorithmic intelligence

There is no established unifying theory or model that guides research into algorithmic intelligence. The main theory is the Computational Theory of Mind, which rests on the presumption that human intelligence is or can be effectively simulated as a software program running on a device. As pointed out by Professor Johnston, two strategies are used to determine how humans acquire, process, retain and use knowledge as a basis for action or for generating further knowledge[19]:

> (i) To build software models of cognitive processes, such as decision making, and then refine them by comparing the performance of the software code to the behaviour of humans under similar conditions. This is called the 'dry cognitive science strategy'. That algorithmic intelligence cannot be achieved without consciousness and feelings is ignored by the dry cognitive science strategy.
> (ii) Consider the brain as a general purpose computer by studying the effects of electrical or chemical simulation on real brains, examine the effect of brain damage, or to attempt to understand what is going on in the brain when observing individuals

[14] Kevin Warwick, *Artificial Intelligence: The Basics* (Abingdon: Routledge Taylor & Francis Group, 2012), p.5.
[15] Amrita Chandradas, "My First Robot", *Financial Times Magazine*, 15/16 July 2017, pp.26–31, at p.30.
[16] Ulric Neisser, "The Imitation of Many by Machine" (1963) 139(3551) *Science* 193, 195, cited in Herbert A. Simon, "Motivational and Emotional Controls of Cognition" (1967) 74(1) *Psychological Review* 29, in which Simon discusses how motivational and emotional controls over cognition can be incorporated into an information processing system.
[17] See Martyn Thomas, "Artificial Intelligence", lecture given at the Museum of London on 13 June 2017, *https://www.gresham.ac.uk/lectures-and-events/artificial-intelligence* [Accessed 29 September 2017]; and Alan Turing, "Computing Machinery and Intelligence" (1950) LIX(236) *Mind* 433 for a discussion on this point and other objections to machine consciousness; John Searle, "Minds, Brains, and Programs" (1980) 3 *Behavioral and Brain Sciences* 417, and John Searle, *Minds, Brains, and Science* (Cambridge, MA: Harvard University Press, 1984) for the Chinese room argument.
[18] Todd Bishop, "Q&A: 'Final Jeopardy' author Stephen Baker on Watson, IBM, the future of e-books and the fate of humanity" (17 May 2011), *GeekWire*, *https://www.geekwire.com/2011/qa-final-jeopardy-author-stephen-baker-watson-ken-jennings-future-humanity/* [Accessed 29 September 2017]; John Searle, "Watson Doesn't Know It Won on 'Jeopardy!'", *Wall Street Journal*, 23 February 2011 (subscription required) made the same conclusion.
[19] Victor S. Johnston, *Why We Feel: The Science of Human Emotions* (New York: Perseus Books, 1999), pp.3–9.

engaged in information-processing tasks. This approach is called 'wet cognitive science'.

These two approaches lead to conflicting views on the nature of our mental states and their relationship to the physical world. More recently, the Computational Theory of Mind has come under pressure from various rival paradigms.

Another theory is reverse-engineering. That is, to construct an intelligent system from information obtained from a detailed investigation of the brain, mainly using scanning technologies. The problem with this is that structure is not sufficient to provide meaning.[20] Neurons are observed moving signals around, but we do not know the purpose of the signals. To reverse-engineer, you need to know the underlying principles. It is a circular exercise. You have to know about what you are looking for in order to recognise it.

Professor Russell and Peter Norvig have set out a number of definitions of algorithmic intelligence.[21] The two major distinctions in defining algorithmic intelligence are thought process and reasoning, and behaviour. They describe four approaches[22]:

(i) Thinking as a human, or the Turing approach, where success is measured in terms of fidelity to human performance. The software has to have, at least, the following capabilities: natural language processing; knowledge representation; automated reasoning and machine learning. This list is not exhaustive.

(ii) Thinking humanly, or the cognitive modelling approach, where success is measured against an ideal performance measure. This is predicated on understanding how the mind works, including attributes such as introspection, psychological experiments and brain imaging.

(iii) Thinking rationally, or the 'laws of thought' approach, which is from the irrefutable reasoning processes and the logicist tradition, where the emphasis is on correctness of inferences. The obstacles are that it is difficult to take informal knowledge and state it in formal terms for logical notation, especially when knowledge is less than 100 per cent certain – which is most of the time – and there is a difference between solving a problem in principle and in practice. To deal with this, the late Professor Simon introduced the concept of 'good enough' or 'satisfice', and noted that the behaviour of an artificial system may be influenced by its ability to adapt.[23]

(iv) Acting rationally, or the rational agent approach, which is where an agent acts – as does software code, but the intention is for the software code to do more, that is to act autonomously – to act so as to achieve the best outcome, or where there is uncertainty, the best expected outcome (this begs the question as to whom determines how to achieve the best outcome and by what criteria).

The debate has ranged between two concepts, that of "weak algorithmic intelligence", dealing with the assertion that software code could act as if it were intelligent, which is the most popular; and "strong algorithmic intelligence", dealing with the assertion that software code is actually thinking. Professor Partridge also notes that Ray Kurtzweil introduced a new category called "narrow" algorithmic intelligence. That is, writing software code that can perform a particular set of functions that used to require human intelligence.[24] This is possibly what Professor Russell and Peter Norvig refer to when providing examples of algorithmic intelligence systems that exist today,[25] such as "robotic" vehicles; speech recognition; autonomous planning and scheduling[26]; playing games; dealing with spam; logistics planning; robotics and machine translation. This emphasises the fact that smaller problems provide more achievable goals and there are an ever-increasing number of positive results. Professor Bostrom has probably given the game away by discussing "superintelligence", which implies that "algorithmic intelligence", as such, is with us now.[27]

Testing and evaluating algorithmic intelligence

A number of tests have been proposed to determine whether an inert machine is capable of demonstrating algorithmic intelligence. The first and best known is the test proposed by Alan Turing, called the Imitation Game,

[20] Partridge, *What Makes You Clever?* (2014), p.19.
[21] Russell and Norvig, *Artificial Intelligence* (2016), p.1.
[22] Russell and Norvig, *Artificial Intelligence* (2016), pp.2–5.
[23] Herbert A. Simon, *The Sciences of the Artificial*, 3rd edn (Cambridge, MA: MIT Press, 1996), pp.28–29; 119–121.
[24] Partridge, *What Makes You Clever?* (2014), pp.181–182.
[25] Russell and Norvig, *Artificial Intelligence* (2016), pp.28–29.
[26] The port of Rotterdam is the first port in the world to use automated guided vehicles and automated terminals: *https://www.portofrotterdam.com* [Accessed 29 September 2017].
[27] Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014). Confusingly, Professor Bostrom indicates that inert software code in the form of superintelligence might bring with it existential catastrophe or destruction of humans. It will be interesting to know whether those living on the island of Run, famous for its nutmeg, who hardly use any technology, will be affected if such were ever to occur (pp.140; 181–182). No doubt the failure to stop killing whales and other land species, and not dealing with climate change sufficiently adequately, will mean that this is never tested. A more realistic possibility is that the failure of public policy will mean that humans will "drift into a position of such dependence on the machines that it would have no practical choice but to accept all of the machines' decisions", as noted by Bill Joy in his essay "Why the future doesn't need us" (1 April 2000), *Wired*, *https://www.wired.com/2000/04/joy-2/* [Accessed 29 September 2017].

now known as the Turing test. This is a behavioural intelligence test. The test is whether software code, by responding to questions and generally entering a "conversation" initiated by a human being, can persuade the human being that they are engaging with another human being, thereby apparently demonstrating that the way the software code performs can be considered to be "thinking" and therefore "intelligent". Professor Partridge set out the underlying assumption[28]:

> "The underlying assumption of the Turing Test is: if you can hold up your end of a prolonged chat in English, then you must be thinking. This means that you are intelligent. Many scientists hold this to be so. Is it reasonable?"

Professor Johnston puts the argument in a slightly different way[29]:

> "Imagine applying the Turing test to a ventriloquist's dummy. Certainly the dummy could answer all your questions in a very human manner. Your conclusion, however, would be, not that the dummy thinks, but simply that the dummy has been programmed by someone who thinks. Isn't this all that can be concluded from passing the Turing test?"

Perhaps the Turing test is more correctly a test that demonstrates that humans can be deceived—as we can be deceived by other people,[30] or of machine human-ness, not necessarily a measure of algorithmic intelligence, or further, as Jaron Lanier puts it[31]:

> "What the test really tells us, however, even if it's not necessarily what Turing hoped it would say, is that machine intelligence can only be known in a relative sense, in the eyes of a human beholder."

It is a very difficult challenge, and to the present, all attempts have failed to pass the test. For instance, at an event organised by the University of Reading held at the Royal Society in London on 6 and 7 June 2104,[32] a computer program named "Eugene Goostman", developed by Vladimir Veselov and Eugene Demchenko, convinced 33 per cent of the human judges that the software program was *imitating* a *simulation* of a 13-year old boy from Ukraine during a Turing test.[33] This event was marketed as a success,[34] and was widely reported by specialist science reporters in the media as such. This was another example of the uncritical assertion of accomplishment, which Professor Partridge calls "hopeware"[35] and Andrew Smart calls "vapourware".[36] It should be noted that the *intelligence* is with the authors of the software code.

Other tests include the Loebner prize[37]; the Feigenbaum test[38]; the Minimum Intelligent Signal Test[39]; and the Ebert test.[40]

The "success" of algorithmic intelligence

There is no doubt that software code has succeeded to such an extent that humans would find it excruciatingly difficult to live if the complex software systems presently in place were to stop working. Life would, literally, stop in many countries until emergency measures were put in place. In the context of algorithmic intelligence, it is possible to agree with Professor Russell and Peter Norvig that there have been successes in the implementation of software code within the vague definition of "weak" algorithmic intelligence.

Games where a human is beaten by software code are often cited as an illustration of the application of algorithmic intelligence. However, those that make this link hide the leap in logic they are making, that algorithmic intelligence is better adapted to static or slowly changing domains—in the words of Cathy O'Neil:

> "The current technology is exceptionally good at navigating finite universes with well-defined rules and principles. That's why chess, checkers, poker, or go champions can be challenged or even beaten by algorithms."[41]

[28] Partridge, *What Makes you Clever?* (2014), p.4.
[29] Johnston, *Why We Feel* (1999), pp.57–58.
[30] Robert Newman, *Neuropolis: A Brian Science Survival Guide* (New York: William Collins, 2017), p.59.
[31] Lanier, *You Are Not A Gadget* (2011), p.31.
[32] Kevin Warwick and Huma Shah, "Can machines think? A report on Turing test experiments at the Royal Society" (2015) 28(6) *Journal of Experimental & Theoretical Artificial Intelligence* 989; for a more realistic appraisal, see Matthew Sparkes, "Computer passes the Turing Test? I'm not convinced" (9 June 2014), *The Telegraph*, *http://www.telegraph.co.uk/technology/news/10886389/Computer-passes-the-Turing-Test-Im-not-convinced.html*; "No, A 'Supercomputer' Did NOT Pass The Turing Test For The First Time And Everyone Should Know Better" (9 June 2014), *techdirt*, *https://www.techdirt.com/articles/20140609/07284327524/no-supercomputer-did-not-pass-turing-test-first-time-everyone-should-know-better.shtml*; in 2011, Cleverbot was voted 59.3% human in India, for which see Jacob Aron, "Software tricks people into thinking it is human" 6 September 2011) *New Scientist*, *https://www.newscientist.com/article/dn20865-software-tricks-people-into-thinking-it-is-human/?DCMP=OTC-rss&nsref=online-news* [All accessed 29 September 2017].
[33] Ray Kurtzweil has a long-term wager with Mitch Kapor in which Ray Kurtzweil predicted that a computer program would pass the Turing test by 2029, and Mitch Kapor has predicted that this would not happen. The bet is for US$20,000 to be donated to the charity of the winner's choice. Payment has yet to be made. See a discussion in "Response to announcement of chatbot Eugene Goostman passing the Turing test" (10 June 2014), *http://www.kurzweilai.net/ask-ray-response-to-announcement-of-chatbot-eugene-goostman-passing-the-turing-test* [Accessed 29 September 2017].
[34] See *http://www.reading.ac.uk/news-and-events/releases/PR583836.aspx* [Accessed 29 September 2017].
[35] Partridge, *What Makes you Clever* (2014), Ch.15.
[36] Andrew Smart, *Beyond Zero and One: Machines, Psychedelics, and Consciousness* (London: OR Books, 2015), p.17.
[37] See *http://www.loebner.net/Prizef/loebner-prize.html* [Accessed 29 September 2017].
[38] Edward A. Feigenbaum, "Some challenges and grand challenges for computational intelligence" (2003) 50(1) *Journal of the ACM* 32.
[39] Chris McKinstry, "Minimum Intelligent Signal Test: An Alternative Turing Test", *http://hps.elte.hu/~gk/Loebner/kcm9512.htm* [Accessed 29 September 2017].
[40] Jennifer Lee, "Roger Ebert Tests His Vocal Cords, and Comedic Delivery" (7 March 2011), *New York Times*, *http://www.nytimes.com/by/jennifer-8-lee* [Accessed 29 September 2017].
[41] Cathy O'Neil, *Weapons of Math Destruction How Big Data Increases Inequality and Threatens Democracy*, reprint (New York: Broadway Books, 2017), p.228.

Indeed, Richard Susskind and Daniel Susskind have asserted that "Deep Blue" was a landmark development of algorithmic intelligence.[42] To understand why this is not correct, it is necessary to analyse the facts carefully, and perhaps to review the failure to follow-up any progress. An IBM software program named Deep Blue beat Garry Kasparov in 1997. The first point to note is that Kasparov knew he was playing against software code and not a human being, as he indicated[43]:

> "I have to tell you that it's a very unusual challenge because since the age of 12 or 13 before any serious match, I have had an opportunity to look at least a few games of my opponents.
>
> I could study my opponent, I could draw a game plan, a war plan and prepare certain surprises. Today I have to shoot in complete darkness."

In the first game, the instructions caused the software code to initiate a move that seemed completely pointless. Kasparov wondered why, but the instructions in the software program caused it to resign. Nate Silver explained[44]:

> "Toward the end of my interview with him, [Murray] Campbell somewhat mischievously referred to an incident that had occurred toward the end of the first game in their 1997 match with Kasparov.
>
> 'A bug occurred in the game and it may have made Kasparov misunderstand the capabilities of Deep Blue', Campbell told me. 'He didn't come up with the theory that the move it played was a bug.'
>
> The bug had arisen on the forty-fourth move of their first game against Kasparov; unable to select a move, the program had defaulted to a last-resort fail-safe in which it picked a play completely at random. The bug had been inconsequential, coming late in the game in a position that had already been lost; Campbell and team repaired it the next day. 'We had seen it once before, in a test game played earlier in 1997, and thought that it was fixed', he told me. 'Unfortunately there was one case that we had missed.'
>
> In fact, the bug was anything but unfortunate for Deep Blue: it was likely what allowed the computer to beat Kasparov. In the popular recounting of Kasparov's match against Deep Blue, it was the second game in which his problems originated—when he had made the almost unprecedented error of forfeiting a position that he could probably have drawn. But what had inspired Kasparov to commit this mistake? His anxiety over Deep Blue's forty-fourth move in the first game—the move in which the computer had moved its rook for no apparent purpose. Kasparov had concluded that the counterintuitive play must be a sign of superior intelligence. He had never considered that it was simply a bug."

In the second game, Kasparov resigned after the instructions written in the software code caused it to make a move that Kasparov thought the software made because the software had the capability to foresee 20 or more moves beyond this point. It was his assumption. This was a mistake by Kasparov. He was so nonplussed by the odd software move in the first game that he did not think to explain this move as a possible error in the code.[45] In the final game, Kasparov opened with an opening that he was not familiar with. He lost to the software code. But he was exhausted. After the first series of games in 1996, the processing power of the computer was doubled, the heuristics refined, and the software was specifically written to beat Kasparov.[46] In addition to which, the team was permitted to alter the program between rounds.[47] This meant that the error noted above was fixed the next day. In later games with software code, no code modifications were permitted between games.[48] This was not an example of algorithmic intelligence. It was an example of a team of software programmers that had worked out how to represent moves in a game of chess. The success was with the programmers, not algorithmic intelligence.

There are many other examples of programs that are designed for a specific purpose, and invariably become the subject of unrealistic hysteria. For instance, the IBM software program named "Watson" was programmed to answer precisely the type of questions posed in a game called *Jeopardy*.[49] Yet it must be emphasised that this exercise was conducted in a way that gave the software code an advantage—the questions were asked in text, and the software code was not required to answer questions by understanding the human voice. Matters would have

[42] Richard Susskind and Daniel Susskind, *The Future of the Professions: How Technology Will Transform the Work of Human Experts* (Oxford: Oxford University Press, 2015), p.164. This observation does not detract from the overall tenor of their message. See Book Reports (2016) 13 *Digital Evidence and Electronic Signatures Law Review* 229.

[43] David Goodman and Raymond Keene, *Man versus Machine: Kasparov versus Deep Blue* (Cambridge, MA: H3 Publications, 1997), pp.48–49; Garry Kasparov, "The Chess Master and the Computer" (11 February 2010), *New York Review of Books*, *http://www.nybooks.com/articles/2010/02/11/the-chess-master-and-the-computer/.*. [Accessed 29 September 2017].

[44] Nate Silver, *The Signal and the Noise* (London: Penguin, 2012), p.288; see also pp.276–279; Goodman and Keene, *Man versus Machine* (1997), described a number of moves by the software code in the first game as being "weird", a "strategic blunder" (p.59), a "counsel of despair" (p.62) and an "inferior choice" of move at the end game (p.63); and Chung-Jen Tan, manager and spokesman for the Deep Blue project, said the next day, in answer to a question, that "We let it have a couple of cocktails" (p.65).

[45] Goodman and Keene, *Man versus Machine* (1997), p.69. In the discussion after the final game, Kasparov said, at p.110, "I have to tell you that game two had dramatic consequences and I never recovered after this game. Not because I lost this game. In fact, I could have made a draw instead of resigning."

[46] Murray Campbell, A. Joseph Hoane Jr and Feng-hsiung Hsu, "Deep Blue" (2002) 134 *Artificial Intelligence* 57; Feng-hsiung Hsu also wrote a book, *Behind Deep Blue: Building the Computer that Defeated the World Chess Champion*, with a New Afterword by the author (Princeton, NJ: Princeton University Press, 2004); and see more recently, Garry Kasparov and Mig Greengard, *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins* (London; John Murray, 2017).

[47] Goodman and Keene, *Man versus Machine* (1997), p.55; Silver, *The Signal and the Noise* (2012), p.284.

[48] Magnus Carlsen, the current world chess champion, has declared that he will never play a software game at chess: Muhir Bose, "Grandmaster's next move", *Financial Times, Weekend*, 8/9 October 2016, House and Home, p.2.

[49] Partridge, *What Makes You Clever?* (2014), pp.381–395; David Ferrucci, Eric Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya A. Kalyanpur, Adam Lally, J. William Murdock, Eric Nyberg, John Prager, Nico Schlaefer, and Chris Welty, "Building Watson: An Overview of the DeepQA Project", *AI Magazine*, Fall, 2010, pp.59–79.

been different had the machine been required to answer questions that were put verbally, in the words of Stephen Baker[50]:

> "Oh, it would have been disastrous, because Watson needed the three seconds between the time it received the text and the time Alex Trebek finished reading it, to process. And it would need another second or two to go through all of its understandings of Alex Trebek's sentence to even come up with its understanding before it began its hunt. So it would be hopeless.
>
> But Ken Jennings and Brad Rutter also read, think and buzz. Alex Trebek's voice is just background noise. They're only listening to him to say, 'When can I press the buzzer?' Because they've read the clue. They're reading just like Watson. Except they have the advantage that they read like human who understand English as a native language, and Watson is struggling with things that humans would never struggle with."

The Twitter "chatbot" Tay by Microsoft was supposed to be an algorithmic intelligence chatterbot, which was placed online on 23 March 2016. Within hours, human beings used deliberately offensive language to "teach" the software, and it was taken offline within 24 hours. It is possible that the humans who wrote the software code did not include code relating to inappropriate behaviour (which, if included, means that the humans writing the code were limiting the reaction of the code), or if it had, the code was seriously lacking. Finally, a lawyer chatbot dealing with parking fines has recently been praised for being an example of "algorithmic intelligence", which it is not. All the software program does is work out whether an appeal is possible through a series of simple questions, such as whether the parking signs were clearly visible, and then guides users through the appeals process.[51]

To counterbalance the apparent negative observations noted above, there is no doubt that software code has improved the lives of many people, including the provision of cochlear implants[52]; the use of virtual reality that helps people that have had a stroke to use weakened limbs[53]; helping make bionic hands work[54]; dealing with epilepsy[55]; possibly helping children with autism[56]; and removing human bias when allocating work,[57] among many other examples that could be cited.

The algorithmic intelligence movement continues apace, and some ventures remain somewhat vague, as described by George Dyson.[58] Dyson was invited to the headquarters of Google in California in October 2005 on the occasion of the 60th anniversary of Von Newmann's proposal to Lewis Strauss for the MANIAC. Google were attempting to execute the strategy that Turing had in mind. That is, gathering all available answers, inviting all possible answers (to what is not clear), and mapping the results. The project had begun to scan all the books in the world. Book lovers objected because the books might lose their souls. Dyson continued[59]:

> "Are we scanning the books and leaving behind the souls? Or are we scanning the souls and leaving behind the books?
>
> 'We are not scanning all those books to be read by people', an engineer revealed to me after lunch. 'We are scanning them to be read by AI'."

This response is an illustration of what Lanier refers to as "cybernetic totalism",[60] and implies that the person making the remark fails, as Dyson points out, to comprehend that reading does not imply understanding, or lead to understanding. And the acquisition of knowledge does not mean that enlightenment or wisdom will follow—as James Gleick points out, "information is not knowledge, and knowledge is not wisdom".[61] Professor Weizenbaum set out the problem in 1976[62]:

> "The capacity of the human mind for sloppy thinking and for rationalizing, for explaining away the consequences of its sloppy thinking, is very large."

Nate Silver commented that if a person

> "thinks of the computer as a sentient being, or the model has having a mind of its own – it may be a sign that there isn't much thinking going on at all … [But] computers are themselves a reflection of

[50] Todd Bishop, "Q&A: 'Final Jeopardy' author Stephen Baker on Watson, IBM, the future of e-books and the fate of humanity" (17 May 2011), *GeekWire*, *https://www.geekwire.com/2011/qa-final-jeopardy-author-stephen-baker-watson-ken-jennings-future-humanity/* [Accessed 29 September 2017]; see also Stephen Baker, *Final Jeopardy: The Story of Watson, the Computer That Will Transform Our World* (Boston/New York: Mariner Books, 2012).
[51] Samuel Gibbs, "Chatbot lawyer overturns 160,000 parking tickets in London and New York" (28 June 2016) *The Guardian*, *https://www.theguardian.com/technology/2016/jun/28/chatbot-ai-lawyer-donotpay-parking-tickets-london-new-york* [Accessed 29 September 2017].
[52] Michael Chorost, *Rebuilt: How Becoming Part Computer Made Me More Human* (New York: Mariner Books, 2005).
[53] "Virtual reality helps stroke patients use weakened limbs" (9 June 2015), *BBC News*, *http://www.bbc.co.uk/news/health-33050784* [Accessed 29 September 2017].
[54] "Josh Cathcart, nine, becomes youngest bionic hand wearer" (21 August 2015), *BBC News*, *http://www.bbc.co.uk/news/uk-scotland-edinburgh-east-fife-34018237* [Accessed 29 September 2017].
[55] "New robot device which may have 'cured' one teenager's epilepsy" (5 February 2016), *http://www.independencehomes.co.uk/news-events/news/item/236-new-robot-device-which-may-have-cured-one-teenagers-epilepsy*.
[56] "Humanoid robot to help children with autism to be trialled in the NHS" (3 April 2017), *http://www.nihr.ac.uk/news/humanoid-robot-to-help-children-with-autism-to-be-trialled-in-the-nhs/6017* [Accessed 29 September 2017].
[57] Sarah O'Connor, "When your boss is an algorithm", *Financial Times Magazine*, 10/11 September 2016, pp.12–18 at p.16: a company called Percolata Inc uses software to rank employees. The software code appears to enable successful employees to be given more work, rather than be at the mercy of their relationship with a manager. Be in no doubt that this is a rare use of software code that appears to be written in the interests of ordinary workers. For the pernicious effects that software code has on teachers and low-paid employees (and the knock-on effect it has on the children of low-paid employees), see O'Neil, *Weapons of Math Destruction* (2017), especially Chs 6 and 7.
[58] George Dyson, *Turing's Cathedral: The Origins of the Digital Universe* (London: Penguin, 2012).
[59] Dyson, *Turing's Cathedral* (2012), pp.312–313.
[60] Lanier, *You Are Not A Gadget* (2011), p.24.
[61] James Gleick, *The Information: A History, a Theory, a Flood* (London: Fourth Estate, 2011), p.409.
[62] Joseph Weizenbaum, *Computer Power and Human Reason* (San Francisco: W.H. Freeman and Co, 1976), p.35.

> human progress and human ingenuity: it is not really 'artificial' intelligence if a human designed the artifice".[63]

David Lubar has expressed the point in a slightly different way[64]:

> "Can machines think? Let's ask the computer. It will know. Actually, it doesn't really matter if computers can think. What matters is whether we think they can think. Of course, if they think that we think that they think, then we might think differently about them."

In conclusion, it might be considered to be an extreme position to assert that if software code cannot explain the "why" of answers, then the answers are not intelligent.[65] Arguably, no inert machinery controlled by software code will be intelligent—but that does not mean that the algorithms cannot be accurate and correct. This is demonstrated in a medical program called "Deep Patient", which is able to predict the beginning of psychiatric disorders such as schizophrenia—something that is particularly difficult for doctors to predict—but nobody knows how the software does this.[66] This illustrates that it does not necessarily matter that a system can present conclusions without its being possible to state the reasons for reaching the conclusions, because the reasons a doctor can present for a diagnosis may be no more than the rationalisation of an intuition, and because a diagnosis should probably be seen as hypothesis rather than a conclusion. At other times it will be important to understand the reasons for a conclusion, although it does not follow that all decisions have been, are or will be made after carefully weighing the facts and producing reasons to justify the conclusion. Reasoning after the event is always a possibility.

The response of judges, lawyers and legal academics

With the increasing use of machine learning systems, it becomes even more incumbent on the law to require evidence of how software systems work, given that the system itself is not capable of offering an explanation for a decision, and the complexity is such that their programmers cannot understand such systems.[67] This part of the article illustrates the problems and sets out the issues that ought to be considered.

Two accidents involving what are often described as "intelligent autonomous motor vehicles" serve to illustrate the extent of the problems the law faces now, not in the future.[68] The first incident occurred on 14 February 2016, involving a Lexus SUV under the control of software code. The vehicle was driving along El Camino Real in Mountain View. The software moved the direction of the vehicle to the far right lane to make a right turn on to Castro Street, but caused the vehicle to stop when sand bags were detected that blocked the direction of travel. The sand bags were positioned around a storm drain. It was the move to get around the sand bags that caused the trouble, according to the report[69]:

> "After a few cars had passed, the Google AV began to proceed back into the center of the lane to pass the sand bags. A public transit bus was approaching from behind. The Google AV test driver saw the bus approaching in the left side mirror but believed the bus would stop or slow to allow the Google AV to continue. Approximately three seconds later, as the Google AV was reentering the center of the lane it made contact with the side of the bus."

Google's car was in autonomous mode and driving at 2 m.p.h. at the time of the crash. The bus was driving at about 15 mph. No injuries were reported, but the front left wheel and fender (the part of the body of the vehicle that frames the wheel well) of Google's car were damaged.

The second incident occurred with a Tesla model S motor vehicle at about 4:40 p.m. on Saturday 7 May 2016. The preliminary report sets out what happened[70]:

> "About 4:40 p.m. eastern daylight time on Saturday, May 7, 2016, a 2015 Tesla Model S, traveling eastbound on US Highway 27A (US-27A), west of Williston, Florida, struck and passed beneath a 2014 Freightliner Cascadia truck-tractor in combination with a 53-foot semitrailer. At the time of the collision, the combination vehicle was making a left turn from westbound US-27A across the two eastbound travel lanes onto NE 140th Court, a local paved road. As a result of the initial impact, the battery disengaged from the electric motors powering the car. After exiting from underneath the semitrailer, the car coasted at a shallow angle off the right side of the roadway, traveled approximately 297 feet, and then collided with a utility pole. The

[63] Silver, *The Signal and the Noise* (2012), p.293.
[64] David Lubar, *"It's Not a Bug, It's a Feature!" Computer Wit and Wisdom* (Boston: Addison-Wesley Publishing Co, 1995), p.125.
[65] Software code might be able to learn, but it cannot explain what it has learnt, for which see Deep Neural Networks learning Atari 2600 games, *https://deepmind.com/research/dqn/* [Accessed 29 September 2017].
[66] Will Knight, "The Dark Secret at the Heart of AI" (11 April 2017), *MIT Technology Review*, *https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/* [Accessed 29 September 2017].
[67] Royal Society, *Machine Learning: The Power and Promise of Computers that Learn by Example* (April 2017, DES4702), pp.93; 96; 111, Box 5.
[68] The vision system of an "intelligent" autonomous system can fail to identify road signs that have been altered by placing printed stickers over the sign. A standard stop sign can be slightly covered over to cause a vision system to identify it as a sign for a speed limit of 45 mph, for which see Ivan Evtimov, Kevin Eykholt, Earlence Fernandes, Tadayoshi Kohno, Bo Li, Atul Prakash, Amir Rahmati and Dawn Song, "Robust Physical-World Attacks on Machine Learning Models" (7 August 2017), *https://arxiv.org/abs/1707.08945* [Accessed 29 September 2017].
[69] See *https://www.google.com/selfdrivingcar/reports/* [Accessed 29 September 2017].
[70] Preliminary Report, Highway HWY16FH018 (26 July 2016), *http://www.ntsb.gov/investigations/accidentreports/pages/HWY16FH018-preliminary.aspx*; for the final report, see Failure Report, US Department of Transportation National Highway Traffic Safety Administration (PE 16-007, 28 June 2016), *https://static.nhtsa.gov/odi/inv/2016/INCLA-PE16007-7876.PDF* [Both accessed 29 September 2017].

> car broke the pole and traveled an additional 50 feet, during which it rotated counterclockwise and came to rest perpendicular to the highway in the front yard of a private residence. The 40-year-old male driver and sole occupant of the Tesla [Joshua Brown] died as a result of the crash.
>
> …
>
> Tesla system performance data downloaded from the car indicated that vehicle speed just prior to impact was 74 mph. System performance data also revealed that the driver was operating the car using the advanced driver assistance features Traffic-Aware Cruise Control and Autosteer lane keeping assistance. The car was also equipped with automatic emergency braking that is designed to automatically apply the brakes to reduce the severity of or assist in avoiding frontal collisions."

Both of these accidents illustrate significant issues that have faced lawyers and judges for some time. However, the way software code has been dealt with in legal systems has been, in effect, to pass off accidents, deaths and failures in digital systems as a problem with the humans involved, leaving the software code largely untested by cross examination—in other words, failing to properly test causation.

As a side issue, metaphors and the use of the word "robot", which was invented by Karel Čapek in his play *Rossumovi Univerzální Roboti*, are important. Metaphors can be helpful or misleading.[71] Often, because they are superficial, they tend to confuse[72] or disguise the reality, and the use of the term "robot" when discussing software code partly perpetuates the misunderstanding about software code and hardware: that combined, they are merely machines, albeit machines that can be programmed to do some very clever and helpful things. Professor Dennett put it succinctly in an interview with John Thornhill:

> "[H]e argues that it is a 'pernicious fantasy' that is distracting us from a far more pressing technological problem. In particular, he worries about our 'deeply embedded and generous' tendency to attribute far more understanding to intelligent systems then they possess. Giving digital assistants names and cutesy personas worsens the confusion.
>
> 'All we're going to see in our lifetimes are intelligent tools, not colleagues. Don't think of them as colleagues, don't try to make them colleagues and, above all, don't kid yourself that they're colleagues', he says.'"[73]

By anthropomorphising software code,[74] the myth continues that in some way machines, made by humans, cannot be controlled. The implication is that a machine, called a "robot", has mental agency. It is humans that impose their false understanding on the machines they make and call "robots". A robot is a machine controlled by software code.[75]

The response by judges and lawyers

There are a number of significant failings in contemporary legal systems,[76] the first of which only affects a number of common law countries (e.g. Australia, England & Wales; New Zealand, the US):

(i) The presumption that computers are "reliable".[77]

(ii) The process of disclosure (or discovery), where the party that controls the software code refuses to give it up for inspection,[78] together with any other relevant design documents. This is only part of the preparation before trial, because it is necessary to understand the software code, if it is disclosed. Additionally, if the design documents are not forthcoming, the inference must be that the owner or user or programmers do not understand the code.

(iii) The rules that permit business records to be automatically granted admission into legal proceedings without the need for authentication (and by extension, failing to fully test the evidence of corporate employees in evidence).[79]

(iv) Faulty analysis of a problem, where, especially in criminal proceedings, the investigating and prosecuting authorities fail to ascertain the correct position before deciding to charge people.

Each is briefly outlined and discussed below.

[71] Simon, *The Sciences of the Artificial* (1996), p.173.

[72] Patrick O'Brien makes this point in the use of metaphors by the members of the Supreme Court in the case of *R. (on the application of Miller) v Secretary of State for Exiting the European Union (Rev 2)* [2017] UKSC 5; [2017] 2 W.L.R. 583; Patrick O'Brien, "All for Want of a Metaphor: Miller and the Nature of EU Law" (30 January 2017), *UK Const. L. Blog*, *https://ukconstitutionallaw.org/2017/01/30/patrick-obrien-all-for-want-of-a-metaphor-miller-and-the-nature-of-eu-law/* [Accessed 29 September 2017].

[73] John Thornhill, "Lunch with the FT", *FT Weekend*, 4/5 March 2017, p.3.

[74] Matters are made worse when commentators refer to the brain as a "machine": e.g. Colin Blakemore, *The Mind Machine* (London: BBC Books, 1988); Demis Hassabis says that "the brain is the only existence proof we have that a general-purpose experience-based learning system is even possible": "The Mind in the Machine", *Financial Times Magazine*, 22/23 April 2017, pp.13–14 at p.14. The brain is part of a body, which is a living organism.

[75] For which see European Parliament, *Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics* (2015/2103(INL)) (31 May 2016), Committee on Legal Affairs, Rapporteur: Mady Delvaux (Initiative – r.46 of the Rules of Procedure), for which see the criticism and observations by Burkhard Schafer, "Closing *Pandora's Box*? The EU Proposal on the Regulation of Robots" (2016) 19 Pandora's Box 55.

[76] Consideration is given to various types of failure in *The Global Risks Report 2017*, 12th edn (World Economic Forum, 2017), but no consideration is given to the failure of legal systems to understand causation.

[77] See Criminal Justice Act 2003 s.129 (1) and (2); and the discussion in Stephen Mason and Daniel Seng (eds), *Electronic Evidence*, 4th edn (Institute of Advanced Legal Studies for the SAS Humanities Digital Library, School of Advanced Study, University of London, 2017), paras 4.24, 5.11, 5.14, 6.214 to 6.129.

[78] Or where, in England & Wales, the prosecution fails to disclose, for which see HM Crown Prosecution Service Inspectorate and Her Majesty's Inspectorate of Constabulary, *Making It Fair: A Joint Inspection of the Disclosure of Unused Material in Volume Crown Court* (July 2017).

[79] See Criminal Justice Act 2003 s.117 and the discussion in Mason and Seng (eds), *Electronic Evidence* (2017), paras 4.39 to 4.41, 7.131, 7.150.

The "reliability" of computers

The presumption of the "reliability" of computers is a presumption that is not warranted. It is lazy and intellectually dishonest to think that the presumption formulated by the Law Commission in 1997[80]—"In the absence of evidence to the contrary, the courts will presume that mechanical instruments were in order at the material time", and extended to mean "computers are reliable"—is nothing other than a negation of the reality of the world in which we are made to live.[81] This presumption is irrational, not justified (the Law Commission did not provide any evidence to justify the presumption), and acts to make the purveyors of justice blind to the facts of causation, and senior appellate judges to continue, in their ignorance, to insist that machines controlled by software code are reliable. The presumption also illustrates the hypocrisy at the heart of English law. Lawyers write clauses for contracts relating to the use of software code that require the user to accept that the software is not free of errors. Such contract terms are considered so normal that nobody appears to understand this fundamental contradiction between the presumption and the acceptance of flawed software code as being normal. The analysis of a child, as in "Kejserens nye Klæder", demonstrates the paradox[82]:

> "Law Commissioners to judges and lawyers: 'It is correct, computers are in order at the material time.'
>
> Judges: 'Then so be it, computers are presumed to be reliable.'
>
> Child: 'But daddy, if computers are reliable, why do the writers of software contracts always write that the code has errors?'[83]
>
> 'Did you ever hear such innocent nonsense?' said her father.
>
> One person whispered and repeated what the child had said, 'Lawyers write in contracts that software code is not free of errors. The child says the code has errors.'
>
> 'But software has errors, so computers cannot be reliable!' the whole world cried out at last.
>
> And what did the Law Commissioners do?"[84]

This is a clear case of cognitive dissonance. The computer industry is fully aware that software code is full of errors, yet when locked into litigation, commercial organisations will go to extreme lengths to prevent the other side from being given sight of the evidence. The lay person must wonder why judges continue with this schizophrenic attitude. The only sane solution is to acknowledge this error and reverse this presumption immediately.[85]

The need for disclosure or discovery

Judges are notoriously reluctant to order the disclosure or discovery of software code, for which see *R. v Seema Misra*[86] by way of example. The party to whom the request is made invariably argues that their code is proprietary and so confidential, that to release it to the other side, even under a confidentiality agreement authorised by the court, would be unreasonable. That it is unreasonable not to establish causation in legal proceedings when it is necessary to examine software code seems to have become the norm—indeed, the British Government seem to have accepted this principle in Principle 3.4 of *The Key Principles of Vehicle Cyber Security for Connected and Automated Vehicles*, which reads[87]:

> "Organisations ensure their systems are able to support data forensics and the recovery of forensically robust, uniquely identifiable data. This may be used to identify the cause of any cyber, or other, incident."

The unintended acceleration cases in the US illustrate the importance of disclosure of software code, in particular the case of *Bookout v Toyota Motor Corp*, Case No.CJ-2008-7969.[88] The background is set out below (footnotes omitted)[89]:

> "Crucially, in the US *Bookout* case, which was one of the high profile unintended acceleration cases, Selna J ordered the disclosure of the software code. The explanation might be because of two significant, and rather fortuitous, factors. When Jean Bookout was driving her 2005 Toyota Camry, it suddenly accelerated. She took action by pulling the parking brake. By so doing, the right rear tyre left a 100-foot skid mark, and the left tyre left a 50-foot skid. The vehicle continued to speed down a ramp, across the road, and came to rest with its nose in an embankment, injuring her and killing her passenger and best friend Barbara Schwarz. Before she died, Schwarz called her husband and said 'Jean couldn't get her car stopped. The car ran away with us.

[80] Law Commission, *Evidence in Criminal Proceedings: Hearsay and Related Topics* (1997), para.13.13.
[81] For a detailed list of failures and a detailed analysis of this presumption, see Mason and Seng (eds), *Electronic Evidence* (2017), Ch.6.
[82] Hans Christian Andersen, "Kejserens nye Klæder", published in *Den lille havfrue* (C.A. Reitzel, 1837) ("The Emperor's New Clothes", published in *The Little Mermaid*).
[83] Errors in software code are a given in the industry: see *The Economic Impacts of Inadequate Infrastructure for Software Testing Final Report* (May 2002), Prepared for National Institute of Standards and Technology by RTI Health, Social, and Economics Research, *https://www.nist.gov/sites/default/files/documents/director/planning/report02-3.pdf* [Accessed 29 September 2017].
[84] With thanks to the translation at *http://www.andersen.sdu.dk/vaerk/hersholt/TheEmperorsNewClothes_e.html* [Accessed 29 September 2017].
[85] As suggested in Mason and Seng (eds), *Electronic Evidence* (2017), paras 6.228 to 6.229.
[86] The transcript of the case, with an introduction, is published in (2015) 12 *Digital Evidence and Electronic Signature Law Review*, *http://journals.sas.ac.uk/deeslr/article/view/2217* [Accessed 29 September 2017].
[87] Centre for the Protection of National Infrastructure, "The key principles of vehicle cyber security for connected and automated vehicles" (6 August 2017), *https://www.gov.uk/government/publications/principles-of-cyber-security-for-connected-and-automated-vehicles/the-key-principles-of-vehicle-cyber-security-for-connected-and-automated-vehicles* [Accessed 29 September 2017].
[88] The trial was held in the District Court of Oklahoma County State of Oklahoma before the Hon. Patricia G. Parrish, District Judge; Transcript (not proofread) of the trial (14 October 2013) (Reported by Karen Twyford, RPR): examination and cross-examination of Michael Barr, *http://www.safetyresearch.net/Library/Bookout_v_Toyota_Barr_REDACTED.pdf* [Accessed 29 September 2017]
[89] Mason and Seng (eds), *Electronic Evidence* (2017), Ch.6: "The presumption that computers are 'reliable'".

> There's something wrong with the car.' Both the skid marks and the telephone call by Barbara Schwarz undermined any suggestion that the acceleration was due to a physical problem in the cabin of the vehicle."

This is a rare example of a judge ordering a party to provide the software code to the litigant, yet this ought to be standard practice, and it is not. If Selna J had not issued such an order, it is doubtful whether the members of the jury would have found in favour of the plaintiffs.

Business records and authentication

The reluctance of judges to order disclosure is also significant, given the rules that permit business records to be automatically granted admission into legal proceedings without the need for authentication. Such rules are, in the age of software code, out-dated. They refer to a period of time when records were recorded on paper. Now that the vast majority of business records are stored on digital devices controlled by software programs that are full of errors, it cannot be right to continue with permitting organisations—both government and commercial—to benefit from a rule that no longer applies to modern working practices.[90] The attitude of organisations also demonstrates that their assertions, based on their electronic records, need to be the subject of careful scrutiny by the courts, and the evidence of employee witnesses has to be treated with great care.

The example of the diesel emissions scandal serves to illustrate this point.[91] James Robert Liang, a Volkswagen engineer aged 62, entered a plea of guilty on 9 September 2016 for his role in implementing software specifically designed to bypass US emissions tests in Volkswagen "clean diesel" vehicles, and was sentenced on 25 August 2017 by Cox J of the Eastern District of Michigan to 40 months' imprisonment in a federal prison.[92] From about 2006, he and his co-conspirators started to design a new "EA 189" diesel engine for sale in the US, but realised that they could not design a diesel engine that would meet the stricter US emissions standards. They therefore designed and implemented software to recognise whether a vehicle was undergoing standard US emissions testing on a dynamometer or being driven on the road under normal driving conditions (the defeat device), in order to bypass the emissions tests. In May 2008, Liang moved to the US to assist in the launch of VW's new "clean diesel" vehicles in the US market. While working at VW's testing facility in Oxnard, California, his title was Leader of Diesel Competence. He admitted that he helped his co-conspirators continue to lie to the Environmental Protection Agency, California Air Resources Board and VW customers, even after the regulatory agencies started raising questions about the vehicles' on-road performance following an independent study commissioned by the International Council on Clean Transportation, which showed that the diesel vehicles' emissions on the road were up to 40 times higher than shown on the dynamometer.[93] A number of his fellow-conspirators were subsequently charged,[94] including Oliver Schmidt, who was the general manager in charge of the Environment and Engineering Office, located in Auburn Hills, Michigan. He entered a plea of guilty before Cox J on 4 August 2017 to one count of conspiracy to defraud the US, to commit wire fraud and to violate the Clean Air Act; and to one count of violating the Clean Air Act.[95]

At present, only the media appear to effectively challenge such practices, although the internet has given people the opportunity to illustrate the failures of software code publically and collectively in a way that organisations—before the internet—would have ignored, as with "smart" meters that give false readings.[96] To consider a simple example,[97] for some time the insurance sector has encouraged motorists to add a "black box" to their vehicles to monitor their driving. The purpose is to enable the insurer to understand the driving behaviour of the insured, so that the insurer can offer a realistic price for insurance in accordance with the way the insured drives their vehicle. However, media reports illustrate that there are significant problems with such devices, such as the software logging false events, recording the driver as driving at 68 m.p.h. in a 30 m.p.h. zone when the car was stationary, and 63 m.p.h in a 30 m.p.h. zone

[90] Stephen Mason and Allison Stanfield, Ch.7, "Authenticating electronic evidence" in *Electronic Evidence* (2017).
[91] Taken from "Volkswagen Engineer Pleads Guilty for His Role in Conspiracy to Cheat U.S. Emissions Tests" (9 September 2016), *https://www.justice.gov/opa/pr/volkswagen-engineer-pleads-guilty-his-role-conspiracy-cheat-us-emissions-tests*; the indictment is available at *https://www.justice.gov/opa/file/890761/download*; and the r.11 Plea Agreement is available at *https://www.justice.gov/opa/file/890756/download* [All accessed 29 September 2017]. Volkswagen AG entered a plea of guilty to three charges on 10 March 2017 and was sentenced on 21 April 2017.
[92] See *https://www.justice.gov/usao-edmi/pr/volkswagen-engineer-sentenced-his-role-conspiracy-cheat-us-emissions-tests* [Accessed 29 September 2017].
[93] Dr Gregory J. Thompson (Principal Investigator), Daniel K. Carder, Marc C. Besch, Arvind Thiruvengadam and Hemanth K. Kappanna (Co-Principal Investigators), *In-Use Emissions Testing of Light-Duty Diesel Vehicles in the United States, Final Report* (Center for Alternative Fuels, Engines & Emissions, West Virginia University, 15 May 2014), *http://www.theicct.org/sites/default/files/publications/WVU_LDDV_in-use_ICCT_Report_Final_may2014.pdf* [Accessed 29 September 2017].
[94] The complete listed is located at *https://www.justice.gov/opa/pr/volkswagen-ag-agrees-plead-guilty-and-pay-43-billion-criminal-and-civil-penalties-six* [Accessed 29 September 2017].
[95] See *https://www.justice.gov/usao-edmi/pr/volkswagen-senior-manager-pleads-guilty-connection-conspiracy-cheat-us-emissions-tests* [Accessed 29 September 2017].
[96] Laura Lee, "SSE glitch: 'Smart meter said I owed thousands'", *BBC News*, 5 March 2017.
[97] For other examples, see "Tech.View: Cars and software bugs" (6 May 2010), *The Economist*, *http://www.economist.com/blogs/babbage/2010/05/techview_cars_and_software_bugs*; Edward Taylor, "BMW fixes security flaw in its in-car software" (30 January 2015), *Reuters*, *http://www.reuters.com/article/2015/01/30/bmw-cybersecurity-idUSL6N0V92VD20150130*; "Recall Notice: Ford issues safety compliance recall in North America" (2 July 2015), *https://media.ford.com/content/fordmedia/fna/us/en/news/2015/07/02/ford-issues-safety-compliance-recall-in-north-america.html*; John Leyden, "Ford's 400,000-car recall could be the tip of an auto security iceberg" (8 July 2015), *The Register*, *http://www.theregister.co.uk/2015/07/08/ford_car_software_recall_analysis/*; Nora Naughton, "65,844 Land Rovers recalled for doors, sunroofs that can fly open" (9 July 2015), *Automotive News*, *http://www.autonews.com/article/20150709/OEM11/150709872/65844-land-rovers-recalled-for-doors-sunroofs-that-can-fly-open*; John Leyden, "Sixty-five THOUSAND Range Rovers recalled over DOOR software glitch" (14 July 2015), *The Register*, *http://www.theregister.co.uk/2015/07/14/range_rover_recall/*; Iain Thompson, "Toyota recalls 625,000 hybrids: Software bug kills engines dead with THERMAL OVERLOAD" (15 July 2015), *The Register*, *https://www.theregister.co.uk/2015/07/15/toyota_recalls_625000_hybrids_over_enginekilling_software_glitch/* [All accessed 29 September 2017].

when it was on a nearby road[98]; a driver purporting to have driven a motor vehicle at a speed significantly faster than the vehicle was capable, and the software falsely recording the insured as driving continuously for five days, even though the insured removed their box from the vehicle and placed it on a table.[99] Errors in software code can expose the user and innocent bystanders to injury, death and loss (such as loss of cash at an ATM), and such vulnerabilities can also be used to hack into the software of motor vehicles and make changes.[100]

The failure to fully analyse a set of facts

By way of example, the case of *R. v Cahill and Pugh*[101] illustrates the failure of the police and the Crown Prosecution Service to fully understand the claims put to them by the administrators at a hospital. Nurses were alleged to have fabricated blood glucose readings (that is, not having actually taken any readings from patients) and then written them up in paper patient notes. From the perspective of the prosecution, it sees that an offence had clearly been made out. However, the failure to consider a range of relevant issues illustrates a lackadaisical approach to the investigation and decision to prosecute. The prosecution failed to understand the dynamics of software code; the physical glucometers were moved around the wards in the hospital and nobody knew where they were; data taken from patients on the machines failed to enter the central database; the technical staff failed to deal with errors; software code in the devices failed to interact with the central databases on the hospital server, and so on. The trial judge concluded that the prosecution evidence was unreliable and was therefore excluded.[102] The prosecution responded by offering no evidence.[103]

Concluding comments

We now live in a highly complex world. Software code controls our lives to an extent that lawyers and judges do not appear to appreciate, and the evidence is that those responsible for teaching and educating future lawyers also fail to understand this. Software code injures and kills people. Unless there is a change in the legal culture, and the attitude to this important topic is not adequately addressed, people suffering losses, injury and death because of the effect of errors in software code will continue to find it very difficult to find justice.[104] Perhaps we can look to Professor Bostrom and his colleagues to consider this issue, given the purpose of the Future of Humanity Institute, which is as follows:

> "The Future of Humanity Institute's mission is to bring excellent scholarship to bear on big-picture questions for humanity. We seek to focus our work where we can make the greatest positive difference. This means we pursue questions that are (a) critically important for humanity's future, (b) unduly neglected, and (c) for which we have some idea for how to obtain an answer or at least some useful new insight. Through this work, we foster more reflective and responsible ways of dealing with humanity's biggest challenges."[105]

There is no doubt that the problems considered in this article are critically important for humanity's future, and they are unduly neglected by the legal profession.[106] In the absence of significant judicial oversight[107] and a detailed analysis by lawyers of causation that includes software code, the software programmers will undoubtedly have become the wizards of the 21st century,[108] and the Committee of the Machine will make judges and lawyers redundant.[109] It is incumbent on those

[98] Iain Jolly, "Watchdog: Black box fault" (23 May 2012), *Auto Express*, *http://www.autoexpress.co.uk/car-news/consumer-news/35610/watchdog-black-box-fault* [Accessed 29 September 2017].
[99] "Black boxes: Can you trust them to lower your car insurance?" (9 November 2016), *BBC News*, *http://www.bbc.co.uk/news/uk-england-37910773* [Accessed 29 September 2017].
[100] Craig Smith (author of *Car Hacker's Handbook*, 2nd edn (San Francisco: No Starch Press, 2016)) is reported to have said, "You can't Q&A the bugs out … Google hasn't even figured out how to do that." Hannah Kuchler, "Hackers expose holes in road to smarter cars', *Financial Times*, 20/21 August 2016, p.14.
[101] Trial at the Crown Court at Cardiff, before H.H. Judge Crowther QC, October 2015.
[102] *R. v Cahill and Pugh*, the Crown Court at Cardiff, ruling by H.H. Judge Crowther QC, 14 October 2015. The author has been furnished with a copy of this ruling, and it will be published in the 2017 issue of the *Digital Evidence and Electronic Signature Law Review*.
[103] For a detailed assessment of this case, see Stephen Mason, Andrew Sheldon and Hein Dries, Ch.9, "Proof: the technical collection and examination of electronic evidence", in *Electronic Evidence* (2017), in particular "Analysis of a failure" at paras 9.91 to 9.96.
[104] See the case of the late Kaushal Gandhi, who was killed when his car went up to full speed. The media reports do not contain sufficient information to reach a proper conclusion as to what might have occurred. The author wrote to Crispin Butler, the senior coroner at Beaconsfield on 2 February 2017, to be considered a "proper person" (which is within the discretion of the coroner), which would then allow the coroner to provide a copy of the transcript. Mr Butler did not reply to the request. See "Driver's last moments recorded in 999 call as he tells operator car's cruise control 'stuck' at 119mph" (24 November 2016), *The Telegraph*, *http://www.telegraph.co.uk/news/2016/11/24/drivers-last-moments-recorded-999-call-ashe-tells-operator-cars/*; Chris Johnston, "Skoda driver decapitated after claiming car's cruise control was stuck" (24 November 2016), *The Guardian*, *https://www.theguardian.com/business/2016/nov/24/skoda-driver-decapitated-in-stuck-cruise-control-mystery*; Shehab Khan, "Driver decapitated in 119 mph crash after car 'got stuck in cruise control'" (25 November 2016), *The Independent*, *http://www.independent.co.uk/news/uk/home-news/driver-kaushal-gandhi-car-stuck-cruise-control-decapitated-119-mph-crash-a7438451.html* [All accessed 29 September 2017].
[105] For which see *https://www.fhi.ox.ac.uk/research/research-areas/* [Accessed 29 September 2017].
[106] Suggestions to ensure that all lawyers and future lawyers are taught the rudiments of electronic evidence have fallen on sterile ground thus far, for which see Denise H. Wong, "Educating for the future: teaching evidence in the technological age" (2013) 10 *Digital Evidence and Electronic Signature Law Review* 16; and Deveral Capps, "Fitting a quart into a pint pot: the legal curriculum and meeting the requirements of practice" (2013) 10 *Digital Evidence and Electronic Signature Law Review* 23.
[107] For instance, appellate judges in a criminal appeal have accepted the evidence of a shop assistant that used a complex software system to give evidence that the system did not malfunction; such evidence in turn was based on the superstition of touching wood, for which see Mason and Seng (eds), *Electronic Evidence* (2017), para.10.12.
[108] They are already in control, putting the legal response to shame, for which see O'Neil, *Weapons of Math Destruction* (2016). Instead of writing papers on electronic wills (*Making a Will*, Consultation Paper 231 (2017)), for one response to the consultation see *http://ials.sas.ac.uk/about-us/news/law-commission-wills-consultation-draws-expertise-stephen-mason* [Accessed 29 September 2017]), the Law Commission would be better spending taxpayers' money on understanding this massive shift and recommending changing the burden of proof, but it has already rejected a proposal dealing with the presumption that computers are "reliable": Stephen Mason, "Electronic evidence: A proposal to reform the presumption of reliability and hearsay" (2014) 30(1) Computer Law and Security Review 80.
[109] E.M. Forster, "The Machine Stops", a short story published in *Collected Short Stories* (London: Penguin, 1954).

involved with the justice system to understand the world in which we live, but there is little evidence of this occurring to date.